Cross-Sectional Consumption And Inequality: The Information Content Of Household Expenditures

Brant Abbott
Queen's University
abbottbrant@gmail.com

Giovanni Gallipoli
University of British Columbia
giovanni.gallipoli@ubc.ca

Preliminary and incomplete

Abstract

Expenditure data are increasingly used to quantify disparities in resources. One reason is that consumption measures convey information not available in income data. Household expenditures, however, reflect preferences as well as genuine disparities in resources. How informative is the cross-sectional dispersion of consumption about life cycle inequality? We show that the information content of consumption is related to the persistence of household expenditures over short and long horizons. Heterogeneity in preferences or life cycle risk leads to persistence in consumption that cannot be explained by income or wealth data, a phenomenon we call "excess persistence". Estimates of the excess persistence allow us to parse consumption inequality into components due to resources, preferences and risk. We examine two household panels, the U.S. PSID and the Italian SHIW, and document similar patterns of excess persistence in both. Through a quantitative assessment, we gain a better understanding of how much of the variation in preferences, risk, and resources is consistent with our estimates of excess persistence. Roughly 77% of consumption inequality can be accounted for through permanent-income measures based on income, wealth and related data; of the remaining consumption variation, about 17% can be attributed to heterogeneous preferences and 6% is due to private information about idiosyncratic risk that cannot be measured from other data. The latter is about as large as the consumption variation due to age. These findings suggest that caution is needed when using consumption data to assess the impact of redistributive policies. Information from income and wealth data might be preferable when assessing policy impacts on heterogeneous populations.

JEL Classification: E21, D14, D6

Keywords: Consumption, Income, Wealth, Measurement, Inequality, Households, Welfare

1 Introduction

Consumption measures are often viewed as a reliable way to gauge inequality in the distribution of resources. Expenditures reflect the availability of disposable income after accounting for taxes, transfers and possibly unobserved sources of cross-sectional insurance. Moreover, past savings make consumption partly insulated from transitory and noisy income fluctuations (Blundell and Preston, 1998; Blundell, Pistaferri, and Preston, 2008). Nonetheless, there is a growing recognition that household expenditures may differ for reasons unrelated to the distribution of resources, which suggests potential misalignment between consumption and welfare inequality. For example, one household may find it desirable to spend new income as it accrues, leading to periods of low and high consumption, whereas another may choose to save a larger portion out of a similar income flow. It is likely that the savers in the example will exhibit lower consumption than the spenders over extended spells of time. However, a snapshot of cross-household inequality at a point in time would not necessarily be informative about long-term welfare gaps: both households make optimal spending plans that reflect their particular incomes and preferences, and their expenditure gaps might overstate underlying gaps in welfare. In this paper we present new evidence suggesting that the patterns of persistence in expenditures, over short and long horizons, provide valuable information about disparities in economic well being. This finding motivates a quantitative assessment of the welfare implications of observed consumption inequality where we connect the persistence of consumption plans to the resolution of uncertainty (or lack thereof). The main insight of our analysis is that consumption persistence at different horizons conveys useful information about the extent of genuine gaps in resources.

To aid in our analysis, we introduce the concept of *excess persistence* of consumption. To be sure, consumption persistence is not surprising in itself because expenditure decisions depend on income and wealth, which are themselves highly persistent. That expenditure plans are generally persistent is consistent with the observation that household rankings within cohort-specific consumption distributions are fairly stable over time (Jappelli and Pistaferri, 2006). For these reasons, the notion of *excess* persistence must refer to persistence in consumption that occurs above and beyond what we can account for through wealth and income histories, both past and future, or through other variables reflecting differences in permanent-income and cash-in-hand. Excess persistence can only arise, and be measured, when households consistently engage in different consumption choices over a given time horizon and in the face of measurably similar circumstances. Possible explanations for such discrepancies include heterogeneity in preferences or in the distribution of risks and constraints that they face.

We estimate excess persistence by comparing estimates of the conditional and the unconditional expenditure persistence at various time horizons. Through this exercise, we document similar patterns of excess persistence in two separate data sets, the U.S. Panel Study of Income Dynamics and the Italian Survey of Household Income and Wealth. In U.S. data the unconditional elasticity of ten-year-ahead household consumption with respect to current consumption is about 0.58; this number falls to about 0.24 after conditioning on wealth, the ten-year path of income, and many other observable factors. In other words, just about 60% of the correlation between current and future consumption is

explained by the persistence of underlying resources. The remainder is what we denote as excess persistence. For non-durable spending the conditional and unconditional elasticities at a ten-year horizon are smaller (0.45 and 0.18, respectively), which implies that about 60% of the consumption autocorrelation is explained by the persistence of measurable resources. In Italian data, the underlying resources explain about 65%, of the non-durable consumption autocorrelation at a ten year horizon, with the unconditional and conditional elasticities of ten-year-ahead consumption estimated at 0.54 and 0.19, respectively. These numbers are echoed broadly, at different time horizons and with different measures of consumption from either country, suggesting that excess persistence accounts for between 30 and 40% of the cross-sectional variation in the persistent component of household spending.

These observations are possibly related to consumption commitments (see Chetty and Szeidl, 2007, 2016). If it is costly for households to frequently adjust certain categories of consumption in response to income fluctuations, consumption patterns would appear more persistent and they'd approximately aggregate to a habit formation model. Chetty and Szeidl estimate that about half of U.S. household budgets have significant consumption commitments. This could induce excess persistence if a household is unable to adjust spending after a shock, leading to a deviation from what is predicted from their post-shock income. A typical consumption category where adjustments are costly is housing, which exhibits somewhat larger excess persistence than other categories. However, as we document in our empirical analysis, excess persistence extends to a wide array of categories including non-durables (e.g. food) where adjustments can be done more nimbly. Notably, in the case of food, excess persistence is larger than for the aggregate bundle. Part of the excess persistence is likely to occur through the choice of committed expenditures alongside other mechanisms. Taking estimates of excess sensitivity as given, our focus is to quantify how much of the cross-sectional inequality in expenditures is due to household preferences and attitudes (risk, discounting) as opposed to unobserved disparities in the distribution of lifetime resources across otherwise similar households. This distinction is important: if consumption does convey information about inequality that is not readily available in income and wealth measures, this information should be identified separately from variation in idiosyncratic household preferences.

Excess persistence is also related to the so-called consumption puzzles of excess smoothness and excess sensitivity. The latter are well-known features of consumption processes that refer to within household, or aggregate, changes in expenditures and income. In contrast, and by definition, excess persistence refers, to long-run cross-sectional differences; perhaps more importantly, there is no puzzle as to how excess persistence arises because variation in risk, costs and/or inter-temporal preferences would account for it. Rather, the key questions we examine are quantitative in nature: does the cross-sectional distribution of consumption convey unique information about inequality in lifetime resources? Can we use consumption persistence to tease out this information?

To make further inroads and build intuition, a theoretical framework is necessary. We begin by using a modified version of the precautionary saving model in Caballero (1991). This allows for a closed form solution and we use it to show how heterogeneity in risk exposure and saving motives can give rise to excess persistence. Crucially, the model makes clear predictions about the way excess persistence, due to an original overhang of risk, should decay over a given time horizon. The model

builds the key insight that uncertainty and saving motives cumulate (and resolve) over successive time horizons. This process of accumulation and resolution can be heterogeneous across households and shapes their consumption plans. The degree of excess persistence depends on the overhang of unresolved uncertainty and the decay of excess persistence depends on the speed of uncertainty resolution (that is, it depends on how much of the uncertainty within a given time interval is front-loaded). For example, if the risk overhang dissolves fast between periods t and $t + j$, measured excess persistence at horizon j would decrease.

As we show, comparing differences in consumption persistence at short and long horizons provides valuable information about the resolution of uncertainty as well as preference heterogeneity. The validity of the argument rests on whether the excess persistence is, in fact, related to the overhang of consumption risk, self-insurance and preparedness to accommodate future events. To check if this is the case, we test whether economic outcomes during the Covid-19 recession respond to consumption choices made a decade prior. To the extent that Covid was not anticipated, it provides a natural way to assess whether earlier life behaviors, motivated by idiosyncratic circumstances, can account for later life outcomes under an unexpected scenario. If true, this would suggest that Persistently different consumption plans reflect the need to accommodate the overhang of consumption risk and idiosyncratic self-insurance motives.

First, we project PSID consumption in 2005-09 on the same wealth, income and demographic controls used to estimate excess persistence; next, we regress the resulting residuals on an indicator for self-reported financial distress in 2020 during the pandemic. We find that residual expenditures in 2005-09 were significantly higher, on average, for those who experienced financial difficulty in 2020. This indicates that households with higher unexplained expenditures in 2005-09 report more hardship in 2020, after controlling for their ten-year sequence of income, wealth and other characteristics. This is direct evidence that the excess persistence of consumption is related to preparedness to smooth consumption over shocks in future periods. Concurrent evidence can be obtained from wealth data, where higher residual consumption is predictive of lower future wealth.

The Caballero (1991) model is instructive but imposes several restrictions that make it unsuitable for quantitative work. For that, we turn to a richer life-cycle model with heterogeneity in income risk as well as in time and bequest preferences. The model helps to distinguish the viability of different sources of heterogeneity as explanations for excess persistence, and to decompose the consumption distribution into what is determined by resources vis-a-vis preferences. The degree of persistence in the unexplained component of consumption (excess persistence) is an important feature of data that informs how much ex-ante heterogeneity should be included in the model. The key feature of data that allows the preference component of ex-ante heterogeneity to be separated from the income-risk component is the relationship between unexplained consumption and the volatility of later life expenditure. In the data there is a clear negative relationship, meaning that savers face greater consumption risks later in life, despite their larger stocks of wealth, which should only be true if their underlying income risks are greater.

Next, we break down the consumption distribution into its resource-dependent and preference-

dependent components. We find that about 17% of overall consumption inequality can be attrbuted to heterogeneity in time preferences. About 6% of consumption variation is due to households' private information about their idiosyncratic expenditure risk (this variation cannot be easily measured from income and wealth data at the household level and it is roughly as large as the contribution of age variation to consumption dispersion). These findings have important implications for the use of consumption inequality data, especially when assessing the impact of policies that redistribute resources in the economy.

2 A Stylized Model of Excess Persistence

In what follows we illustrate how precautionary behavior, mediated by heterogeneity, leads to excess persistence. We study a variation of the model in Caballero (1991), extended to include heterogeneity in age-varying income risk, and time discounting.

2.1 Model Basics

The model posits time-separable exponential utility (constant absolute risk aversion) with functional form $u(c) = -(1/\gamma)\exp\{-\gamma c\}$. A time-varying discount rate δ_t includes discounting for mortality risk. Income, y_t, follows a unit root process with normally distributed zero-mean innovations, $\{v_t\}$, whose variance is σ_t^2, so that $y_t = y_{t-1} + v_t$. Households hold wealth a_t, which earns a rate of return r. The budget constraint is $a_{t+1} = (1+r)a_t + y_t - c_t$.

Following Caballero (1991, 1990), the optimal consumption rule satisfies

$$E_t[c_{t+1}] = c_t + \lambda_{t+1}, \tag{1}$$

where $\lambda_{t+1} = \frac{\gamma}{2}\sigma_{t+1}^2 + \frac{1}{\gamma}\ln(\frac{1+r}{1+\delta_{t+1}})$ captures the roles of precautionary saving and time-preference in consumption dynamics. Equation (1) can be iterated forward to express $E_t[c_{t+j}] = c_t + \sum_{k=1}^{j}\lambda_{t+k}$. Combining this with the lifetime budget constraint yields a closed-form solution

$$c_t = \frac{r}{1+r}a_t + y_t - \sum_{k=1}^{\infty}\left(\frac{1}{1+r}\right)^k \lambda_{t+k}. \tag{2}$$

This is the canonical permanent-income consumption rule minus the extra term $\sum_{k=1}^{\infty}\left(\frac{1}{1+r}\right)^k\lambda_{t+k}$, which captures non-certainty equivalent savings motives. The solution shows that the cumulated effects of future uncertainties and time-preferences shape contemporaneous consumption.

2.2 From Unresolved Uncertainty to Excess Persistence

To derive excess persistence in the model, we define the vector X_t of observable state variables. Wealth, income and age are included in this vector, i.e. $X_t = (a_t, y_t, t)$; however, in richer models, the vector

can be more expansive. The conditional expectation of c_t given X_t includes permanent income alongside the predictable part of each λ_{t+k}. For example, the age-common components of σ^2_{t+k} and δ_{t+k} are included in $E[\lambda_{t+k}|X_t]$.

We denote the remaining variation as $\tilde{\lambda}_{t+k}(\tilde{\delta}_{t+k}, \tilde{\gamma}_{t+k}, \tilde{\sigma}^2_{t+k})$. These residual terms are a function of heterogeneity in the underlying preference and risk parameters. This dependence is left implicit hereafter.

Consumption at time t can be cast as:

$$c_t = E[c_t|X_t] - \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \tilde{\lambda}_{t+k}. \tag{3}$$

Deviations of consumption from what is predicted from observables depends on the accumulation of idiosyncratic preferences and uncertainty over the entire future. The degree of excess persistence depends on how much of the term $\sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \tilde{\lambda}_{t+k}$ lingers in future expenditures. Put differently, excess persistence captures how much of $\sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \tilde{\lambda}_{t+k}$ persists over a given horizon.

To clarify this point, we proceed in two steps. First, we consider the expression in (2) for period c_{t+j}; second, we compute the optimal wealth dynamics over successive periods as:

$$a_{t+1} = a_t - \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^{k-1} E_t[\Delta y_{t+k}] + (1+r) \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \lambda_{t+k} \tag{4}$$

Substituting the wealth dynamics into the expression for c_{t+j} we obtain the following expression for future consumption c_{t+j}:

$$c_{t+j} = \frac{r}{1+r} a_t + y_t - \left(\sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \lambda_{t+k} - \sum_{k=1}^{j} \lambda_{t+k} \right) + \sum_{k=1}^{j} v_{t+k}. \tag{5}$$

The term $\sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \lambda_{t+k} - \sum_{k=1}^{j} \lambda_{t+k}$ contains the unresolved uncertainty that enters the residual variation $(c_t - E[c_t|X_t])$ in (3). This is the term that generates the excess persistence of consumption between periods t and $t+j$.

The long-term influence of unresolved uncertainty. To see how the unresolved risk overhang influences consumption dynamics at the individual level, we use (3) to rewrite the $j-$ahead consumption as

$$\underbrace{c_{t+j} - E[c_{t+j}|X_t]}_{\text{residual c at t+j}} = \underbrace{\frac{\sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \tilde{\lambda}_{t+k} - \sum_{k=1}^{j} \tilde{\lambda}_{t+k}}{\sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k \tilde{\lambda}_{t+k}}}_{\text{unresolved uncertainty at t+j}} \underbrace{(c_t - E[c_t|X_t])}_{\text{residual c at t}} + \sum_{k=1}^{j} v_{t+k}. \tag{6}$$

The cumulative unresolved uncertainty rescales the variation $(c_t - E[c_t|X_t])$ and determines the persistence of residual consumption between t and $t+j$.

The relationship is positive whenever the present value of cumulated future uncertainty exceeds the raw sum of realized (resolved) uncertainty between periods t and $t+j$. Intuitively, if the largest savings motives lie beyond period $t+j$, then consumption at t and $t+j$ will be similarly affected leading to positive excess persistence. In contrast, if the largest savings concerns at t are resolved prior to $t+j$, then the opposite would occur.

Some special cases to aid intuition. In special cases the coefficient simplifies to reveal the specific forces behind the pass-through of horizon uncertainty into excess persistence. For example, if the idiosyncratic heterogeneity is not age-varying, then $\tilde{\lambda}_{t+j} = \bar{\lambda}$ for all $t+j$. In this case the exact nature of heterogeneity is unimportant and the excess persistence coefficient, that is, the rescaling of $(c_t - E[c_t|X_t])$ in (6), reduces to $(1 - r \times j)$; thus, excess persistence declines linearly at a speed dictated by the financial discount rate.

Another possibility is that heterogeneity in λ_{t+j} wanes over time, perhaps because income risk declines later in life. If we let $\tilde{\lambda}_{t+j+1} = \rho\tilde{\lambda}_{t+j}$ with $\rho \in (0,1)$, the excess persistence coefficient becomes $\left(1 - \frac{1-\rho^{j-1}}{1-\rho}(1+r-\rho)\right)$. In this case excess persistence declines monotonically with j, but does so at a speed that depends on the decay rate ρ. If ρ is close to zero so that cumulated heterogeneity is front-loaded, then the excess persistence coefficient can be negative even for relatively small time windows j.

In more general cases the loading on $(c_t - E[c_t|X_t])$ in (6) can vary across households, and so the excess persistence coefficient from a regression will be related to the cross-sectional average of the loadings. Nonetheless, these regression coefficients convey valuable information about the underlying risk overhang. In a richer quantitative model the intensity of excess persistence can be used to measure heterogeneity in preferences and risk overhang.

3 Evidence on the Excess Persistence of Consumption

3.1 Data Sources

We consider two independent data sources in our empirical analysis: the U.S. Panel Study of Income Dynamics (PSID), and Italy's Survey of Household Income and Wealth (SHIW). Both sources include long-term longitudinal data on households' expenditure, income, wealth and other variables; both have been commonly utilized in the literature.

3.1.1 PSID

Our PSID sample spans the period between 2005 and 2021. To be in the sample, an observation must be directly followed by five waves of complete biennial data, thus the latest observations are those

from 2011 which are followed by complete data up to 2021. Furthermore, household structure must remain intact over that period, i.e. we focus on continuously married couples. These restrictions allow us to study the persistence of consumption at horizons from two to ten years using a consistent set of households. Given nine waves of data, each household may contribute up to four observations to the sample. The 2,097 households in the sample contribute three observations on average, for a total of 6,277 observations.

We choose 2005 as our starting point because in that year a broadened set of expenditure categories was included in the survey. The patterns of the post 2005 data align reasonably well with those in the Consumer Expenditure Survey (Insolera, Simmert, and Johnson, 2021). Households are surveyed on various expenditures within the general categories of food, transportation, clothing, recreation, trips, housing, education, health care and child care.

In our empirical analysis we examine (i) a measure of total spending that includes all items except education, health care and child care, and (ii) a measure which further excludes car payments and housing expenditures in order to emphasize non-durable aspects of spending. To illustrate the robustness of our findings we separately present results for each general category, as well as versions of total spending that include alternative measures of housing expenditures, or that include education, health care and child care spending.

In the analysis we use four measures of income. The first is total after-tax income of the household (calculated using NBER's TAXSIM), and the second and third are the labor incomes of each spouse; the fourth is asset income. Labor incomes enter separately in the analysis as they may contain unique information about the human capital of individuals.

To measure wealth we use net worth including home equity. Asset income is included separately from other income sources, to control for rate of return heterogeneity: we posit that differences in capital income, holding wealth constant, equate to differences in rate of return (for a discussion of measurement issues, see Daminato and Pistaferri, 2023). Other covariates include dummy variables for family size, occupation and age of each spouse, as well as the number of years of education of each spouse.

In the consumption analysis, we allow for non-linear effects of wealth and income by constructing decile-specific dummy variables for each of the income and wealth variables. We also interact age with education, age with earnings, and education with earnings separately for each spouse. It is not uncommon for labor income to be zero, and total family income can sometimes be reported as negative. To accommodate these possibilities, we add \$1,000 to real labor income before applying log transformations, and we apply the neglog transformation to real total family incomes.

In the analysis we control for ten-year leads of the income variables in order to account for information that households may possess about their future resources. For example, an observation from the year 2005 will include income and wealth controls from 2005, 2007, 2009, 2011, 2013 and 2015. The sampling criteria described above ensure that the lead values are available for all observations. Table 1 shows summary statistics for the distributions of expenditures and control variables in our sample.

	PSID Summary Statistics		
	Mean	Standard Dev.	Median
Demographics			
Age of Respondent	46.7	13.3	47
Age of Spouse	44.6	13.0	45
Education Years of Resp.	14.1	2.22	14
Education Years of Spouse	14.1	2.12	14
Household Size	3.1	1.20	3
Spending			
Total	58283.8	39835.6	50946.2
Non-Durable	32718.3	19906.4	27820.1
Housing	25366.7	30542.4	21292.3
Food	9219.8	4607.7	8312.3
Transportation	13445.6	10480.0	10872.7
Clothing	1931.3	2708.8	1091.6
Trips	2615.4	3880.1	1528.0
Recreation	2367.4	3058.2	575.6
Income and Wealth			
Family Income	110902.0	86662.6	92378.2
Respondent's Earnings	59608.0	62260.2	49121.3
Spouse's Earnings	29039.1	30986.2	23825.5
Net Worth	512798.2	2079562.0	162646.0

Table 1: All values are deflated to 2012 dollars using the PCE index.

3.1.2 SHIW

Our treatment of the Italian SHIW data mimics that of the PSID, with some limitations. Like the PSID, the SHIW provides biennial data on consumption, wealth and income. We employ data from 1998 to 2016 in our main analysis. No survey was conducted for 2018 but we supplement this sample with data from 2020 when we study the relationship between excess persistence and household responses to the Covid-19 crisis (an exercise similar to the one we do in the PSID). Our SHIW sample includes all observations of continuously married households for whom ten years (five waves) of complete follow-up data are available. In total, there are 4,286 such observations contributed by 1,646 unique households.

The SHIW reports spending on durables and non-durables separately, but does not provide amounts for disaggregated categories. Therefore, we focus on non-durable spending and total spending, which parallel the two main measures in the PSID sample. Among the control variables, total household income and labor income are available; however, unlike the PSID, we cannot separately control for the earnings of each spouse. We include the same demographic control variables as in the PSID analysis, e.g. age, education and family size. Household net worth is reported in the SHIW, and we allow for non-linear effects of wealth and income through decile dummies, just as in the PSID. Table 1 shows summary statistics on the distributions of expenditures and control variables in the SHIW sample.

	SHIW Summary Statistics		
	Mean	Standard Dev.	Median
Demographics			
Age of Head	52.8	11.8	53
Household Size	3.3	1.06	3
Spending			
Total	31260.2	18176.2	26934.1
Non-Durable	28151.8	14730.4	24800.5
Income and Wealth			
Family Income	36412.2	28222.2	30696.2
Family Earnings	18919.6	18492.7	17222.6
Net Worth	304085.2	466970.6	194045.9
Highest Education of Head	Proportion		
High School Degree	64.0	−−	−−
College Degree or More	26.6	−−	−−

Table 2: All values are deflated to 2015 Euros using the household consumption deflator.

3.2 Measuring Consumption Persistence

To characterize the persistence of household expenditures, we estimate regressions of the form:

$$c_{t+j} = F(X_t) + \beta c_t + \epsilon_{t+j} \tag{7}$$

where c_{t+j} is log-consumption at time $t+j$, X_t includes demographic controls, wealth, and all current and future income variables (observed biennially). The function F accommodates non-linearities and interactions, as described before.

The relationship in (7) captures, through $F(X_t)$, the permanent income components as well as possible non-linearities in the consumption function that arise when certainty equivalence does not hold. As such, (7) is an empirical representation of (6), where β reflects the extent to which variation in c_t that is orthogonal to income and wealth predicts future consumption c_{t+j}. In the absence of excess persistence, we would estimate $\beta = 0$ because persistent differences in household spending would be entirely explained by the evolution of resources. If excess persistence is present, we would estimate $\beta > 0$. As shown in the theoretical analysis, the positive coefficient could arise from various forms of heterogeneity in the parameters of the consumption-savings decision.

Consumption persistence at different time horizons. We begin the analysis begins by estimating a series of regressions that fit (7) to the full sample (all ages combined) over future consumption horizons $j \in \{2, 4, 6, 8, 10\}$. As mentioned before, we account for income and wealth controls for all horizons up to 10 periods ahead. These controls enter regardless of the consumption horizon that we consider.

Each column of Tables 3, 4 and 6 shows estimated, horizon-specific β coefficients for different measures of spending across the two data sets. For the PSID, Table 3 documents considerable excess persistence in consumption: the excess persistence coefficient declines in the horizon j but past consumption exhibits a consistently high degree of added explanatory power over future consumption. This is true also at long lags, even after ten years.

For total spending, c_{t+10} increases by about 0.25% for every 1% increase in current spending, controlling for permanent income. Excess persistence in non-durables is less pronounced: we estimate about a .18% increase in c_{t+10} for every 1% increase in c_t (holding permanent income constant). This seems sensible given the lower commitments often associated with those spending categories. We estimate similar patterns after including education, child care and health care spending with the other components of total spending (forming Total+), or using imputed/actual rent to measure housing expenditures.

Persistence of different consumption categories. Table 4 shows how estimates of excess persistence vary across spending categories. We estimate the highest excess persistence for housing; food and clothing also exhibit more persistence than total or non-durable spending. Other categories,

	Excess Persistence Coefficients (β)			
	PSID Consumption Measure			
Horizon ($t+j$)	Total	Non-Durable	Total+	Total with Rent
$t+2$	0.3827	0.3370	0.3946	0.4183
	(0.0214)	(0.0174)	(0.0212)	(0.0208)
$t+4$	0.3125	0.2552	0.3153	0.3294
	(0.0206)	(0.0168)	(0.0206)	(0.0201)
$t+6$	0.2728	0.2052	0.2724	0.2824
	(0.0207)	(0.0161)	(0.0207)	(0.0196)
$t+8$	0.2567	0.1852	0.2526	0.2676
	(0.0202)	(0.0160)	(0.0204)	(0.0197)
$t+10$	0.2445	0.1823	0.2499	0.2657
	(0.0201)	(0.0164)	(0.0201)	(0.0201)
Sample Size	6457	6457	6457	6457

Table 3: Robust Standard Errors

i.e. transportation, trips or recreation, exhibit marginally lower persistence than total spending, but we continue to find considerable excess persistence in all categories of spending. This suggests that excess persistence is not singularly driven by a particular type of spending, or exclusively by contractual obligations; for example, it is unlikely that food expenditure 10 years out are sticky due to long-term contracting.

Next, Table 6 provides estimates of excess persistence in the SHIW data, again at horizons up to ten years. Despite the differences in measurement, and underlying differences between the countries, the estimates are remarkably similar to the PSID, particularly for non-durable spending. Note that total spending in the SHIW is very different than in the PSID, i.e. durables are only about 10% of total spending in SHIW, as opposed to 44% in the PSID, largely due to the inclusion of housing payments. For non-durables a .144% increase in c_{t+10} is implies for every 1% increase in c_t (holding permanent income constant), as opposed to 0.182% in the PSID. Thus, excess persistence is somewhat less in the SHIW, but still very salient.

Consumption persistence over the life cycle. Does the excess persistence vary over the life cycle? To examine this question, we focus on nondurable spending and employ two approaches. First, we divide our PSID sample into age groups and re-estimate (7) for each group at each horizon. This reveals how patterns of excess persistence change with age without imposing parametric restrictions. Given the smaller sample sizes within age groups, there is some noisiness. Second, we impose parametric restrictions (quadratic variation) on the excess persistence over the life cycle, which results in a smoother stylized summary of the age-variation of excess persistence.

	Excess Persistence Coefficients (β)					
	PSID Results by Category					
Horizon ($t+j$)	Housing	Food	Transport	Clothing	Trips	Recreation
	Version A: c_{t+j} and c_t are category-specific spending					
$t+2$	0.4735	0.4093	0.2318	0.4118	0.3404	0.2832
	(0.0212)	(0.0206)	(0.0169)	(0.0163)	(0.0203)	(0.0159)
$t+4$	0.3940	0.3621	0.1724	0.3812	0.2759	0.2557
	(0.0217)	(0.0185)	(0.0165)	(0.0154)	(0.0190)	(0.0164)
$t+6$	0.3431	0.3525	0.1418	0.3306	0.2426	0.2304
	(0.0194)	(0.0186)	(0.0156)	(0.0159)	(0.0192)	(0.0163)
$t+8$	0.3267	0.3208	0.1390	0.3271	0.2467	0.2567
	(0.0192)	(0.0172)	(0.0155)	(0.0161)	(0.0192)	(0.0168)
$t+10$	0.2990	0.3192	0.1672	0.3254	0.2122	0.2564
	(0.0185)	(0.0174)	(0.0170)	(0.0166)	(0.0200)	(0.0178)
	Version B: c_{t+j} is category-specific spending, c_t is total spending					
$t+2$	0.3827	0.2171	0.2853	0.3855	0.3652	0.4887
	(0.0214)	(0.0195)	(0.0326)	(0.0369)	(0.0458)	(0.0450)
$t+4$	0.3125	0.1776	0.2137	0.3323	0.2971	0.3607
	(0.0206)	(0.0195)	(0.0311)	(0.0375)	(0.0450)	(0.0445)
$t+6$	0.2728	0.1677	0.1716	0.2765	0.2503	0.2978
	(0.0207)	(0.0205)	(0.0312)	(0.0370)	(0.0469)	(0.0443)
$t+8$	0.2567	0.1967	0.1953	0.3176	0.2354	0.2100
	(0.0202)	(0.0203)	(0.0303)	(0.0378)	(0.0448)	(0.0452)
$t+10$	0.2445	0.2011	0.1946	0.3033	0.1907	0.2677
	(0.0201)	(0.0207)	(0.0331)	(0.0394)	(0.0489)	(0.0492)
Sample Size	6457	6264	6164	6037	4228	5051

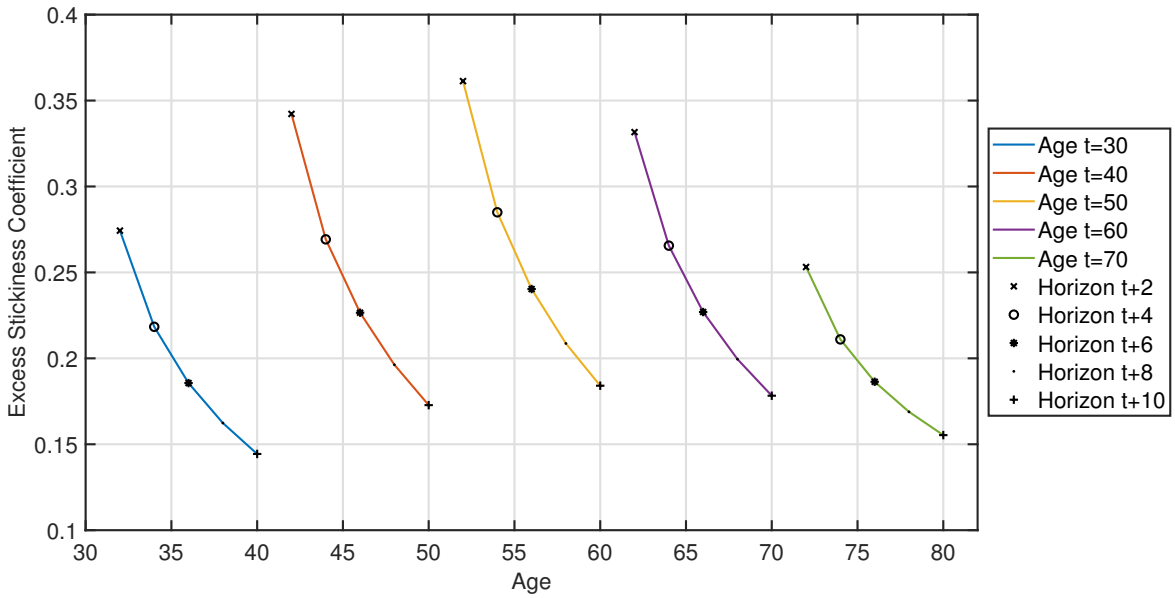
Table 4: Robust standard errors

	Excess Persistence Coefficients (β)	
	SHIW Consumption Measure	
Horizon ($t+j$)	Total	Non-Durable
$t+2$	0.3171	0.3733
	(0.0322)	(0.0346)
$t+4$	0.2764	0.3122
	(0.0308)	(0.0334)
$t+6$	0.2237	0.2886
	(0.0303)	(0.0320)
$t+8$	0.1855	0.2197
	(0.0284)	(0.0318)
$t+10$	0.1582	0.1885
	(0.0281)	(0.0312)
Sample Size	4286	4286

Table 5: Robust standard errors in parentheses.

	PSID Non-Durable Spending			
horizon $t+j$	age$\leq$35	36-47	48-59	age$\geq$60
$t+2$	0.2782	0.3494	0.3900	0.2894
$t+4$	0.2118	0.2657	0.2548	0.2454
$t+6$	0.1568	0.2087	0.2110	0.2292
$t+8$	0.1569	0.2023	0.1941	0.1961
$t+10$	0.1863	0.1714	0.2060	0.1962

Figure 1: Each line is the $j \in \{2, 4, 6, 8, 10\}$ sequence for a particular age. The base ages are 30, 40, ... 70.



3.3 Is Persistence a Response to Uncertainty?

One potential concern is that excess persistence of consumption is the by-product of factors unrelated to consumption uncertainty and self-insurance, such as correlated measurement errors or unobserved behavioral biases. We examine the empirical underpinnings of consumption persistence by performing a few checks.

We begin by testing whether households had different experiences during the Covid-19 recession depending on their consumption behavior a decade or more earlier. Next, we provide ancillary evidence that consumption stickinees is not a spurious relationship. To this purpose, we replace the c_{t+j} in (7) with other outcome variables that are related to heterogeneity in preferences and idiosyncratic consumption risk.

First, we explore how past consumption relates to later life wealth and estimate a strong and negative relationship, which is consistent with heterogeneity in time preferences.

Second, we examine the impact of early life consumption on later life expenditures' volatility; also in this case we find a negative correlation, which indicates that households may have private information about the volatility of their consumption plans: those who face larger risks choose to spend less in order to better self-insure.

Finally, we consider potential heterogeneity in risk attitudes (utility curvature) as reflected in the effect of early consumption on the riskiness of wealth portfolios, measured by stock ownership. We find no evidence of a relationship between early life consumption and later life ownership, suggesting that either risk attitudes have a limited role for excess persistence, or the data we have is not enough to identify such effects.

Past consumption vs economic distress during the Covid-19 pandemic. Do differences in early life consumption and savings change the severity of later life shocks, as in Caballero (1991)? To examine this question, we select data where the initial observation of consumption is from 2005-2009; then, we run regressions of the form in (7) where we replace c_{t+j} with an outcome observed during the 2020 covid period.

In the PSID, we use two alternative outcome variables. The first is an indicator for self-reported financial distress from a special survey related to the pandemic. In total, 9.4% of sample members responded affirmatively to experiencing financial difficulties during that period. The second outcome is the change in household spending between 2018 and 2020. General declines were observed, with overall spending falling by 8.3% on average among sample members; the average spending on non-durables fell by 13.2%. We test whether the fall in consumption was more or less pronounced among those with higher residual spending in the 2005-2009 period.

	Excess Persistence and Pandemic Outcomes	
	Financial Distress	$c_{2020} - c_{2018}$
c_t coefficient	0.0204	-0.0313
	(0.0103)	(0.0137)
Sample Size	4032	3866

Table 6: This table reports coefficients on c_t, measured as non-durable spending, in regressions similar to (7), but with c_{t+j} replaced by the indicated dependent variable. Those with higher residual expenditure in 2005-2009 were significantly more likely to experience financial distress during the Covid-19 pandemic, and experienced significantly larger drops in spending between 2018 and 2020.

Households who chose higher residual expenditure in 2005-2009 were significantly more likely to experience financial distress during the Covid-19 pandemic; they also experienced significantly larger drops in spending between 2018 and 2020. For context, if we had focussed on how expenditure in 2005-2009 unconditionally predicts these outcomes we would find the opposite, with initial high spenders being much less likely to experience financial distress in 2020, which fits the narrative that less affluent households were the most affected by the pandemic.[1] This analysis cannot reject the hypothesis that excess persistence is related to heterogeneity in saving motives. Those who spent less in 2005-2009, relative to similar households, had larger buffer-stocks of wealth in 2020 and were, therefore, less likely to experience financial difficulties.

Early life consumption and target wealth. The Italian SHIW data includes a survey variable that can be interpreted as reflecting the buffer-stock wealth target. For example, Jappelli and Pistaferri (2024) use this variable to study how permanent income shocks affect target wealth, and the manner in which target wealth adjusts over time. Target wealth scales with permanent-income, and so consumption will be highly predictive of target wealth. More subtly, to the extent that the excess persistence of consumption is related to heterogeneity in risk or preferences, consumption residuals may be

[1] A regression of the financial distress indicator on c_t (with no controls) yields a coefficient -0.017 (S.E. 0.008).

negatively correlated with permanent-income residuals (after controlling for permanent-income variables). For example, relatively patient households will choose lower consumption than others with the same permanent-income, and will also have higher buffer-stock targets.

The SHIW survey item reflecting buffer-stock savings is available in only years 2004, 2006, and 2010-2016. The 2004 and 2006 data provide 1761 observations of target wealth in years that coincide with c_t observations. Alternatively, one can consider how c_t and future target wealth option covary, with the 2010-16 data providing 3450 observations at ten-year horizon. We consider both of these possibilities in table 7. In both versions of the regression we see a considerable negative relationship between spending and target wealth after taking out the variation due to permanent-income. A one-percent increase in unexplained spending is associated with a 0.35-0.45 percent decrease in the buffer-stock target, consistent with heterogeneity in time preferences or risk.

	Excess Persistence and Target Wealth	
	Current Target Wealth Residuals	Ten-Year Ahead Target Wealth Residuals
c_t residuals coefficient	-0.3481	-0.4799
	(0.1077)	(0.0799)
Sample Size	1761	3450

Table 7: This table reports coefficients on c_t residuals, i.e. the part of c_t remaining after controlling for $F(X_t)$ as in (7). The dependent variable is residual variation in target wealth, either contemporaneously or ten-years into the future. Those with higher residual expenditure have lower buffer-stock wealth targets, consistent with heterogeneity in time preferences or risk.

Early life consumption and the volatility of later life expenditures. A potentially important source of variation in consumption behaviors is heterogeneity in the income and expenditure risks that households face. Those who face larger risks may save more (spend less) in order to better self-insure against negative outcomes, even though their observable characteristics are the same as other households. To assess for the presence of such private imformation about risk profiles, we examine how the residual spending at time t covaries with the standard deviation of consumption between periods $t+2$ through $t+10$. The first column of Table 8 shows that households who spend more than expected at time t tend to have lower variability in their realized consumption over the next ten years. This result is entirely consistent with the notion that households who face less risk in the future will spend more in the current period. Though the coefficient in Table 8 is not large, at only -0.0115, it is economically meaningful as the average of the outcome variable is 0.22.

Note that the result about consumption volatility is not an obvious one. The variability of future consumption is likely to respond to current savings decisions: for example, if all variation were driven by risk aversion, as opposed to consumption risk itself, we would expect that those who save more experience less consumption variation in the future (a positive relationship). Thus, the negative relationship in the data is a strong indicator that idiosyncratic variation in the risk households face is an

important component of excess persistence. In principle, one might use income volatility, rather than consumption volatility, as the dependent variable; however, the realizations of income between $t+2$ through $t+10$ are already included in the RHS of the regression when we control for the income path over the full horizon; our findings suggests that consumption data relays the private information that households have about their own expenditure risk, above and beyond what is measured by income volatility.

	Excess Persistence and Other Outcomes - PSID			
	Consumption Volatility	Wealth Accumulation	Stock Ownership	Real Estate investment
c_t coefficient	-0.0115	-0.1836	-0.0047	0.0136
	(0.0044)	(0.0778)	(0.0117)	(0.0116)
Sample Size	6457	6457	5847	5847

Table 8: This table reports coefficients on c_t, measured as non-durable spending, in regressions similar to (7), but with c_{t+j} replaced by the indicated dependent variable. Real Estate Investment indicates ownership of real estate that is not the primary home.

Early life consumption and wealth accumulation. Next, we consider whether excess persistence is related to later life wealth holdings. We replace c_{t+j} in (7) with the neglog transform of wealth ten-years into the future (a_{t+10}) and show that, perhaps unsurprisingly, more spending leads to less wealth accumulation. In the full sample, the coefficient on c_t is -0.1835, which is approximately the elasticity of future wealth with respect to current spending. For context, a 1% in non-durable spending is \$310 at the average, while a 0.1835% decrease in ten-year-ahead wealth is \$1,323 at the average. This finding indicates two things: (i) excess persistence is significantly related to saving behavior; (ii) the budget constraints behaves in a reasonable way in the data. The effects of early life consumption on wealth accumulation are similar at short ($j = 2, 4$) and long ($j = 6, 8, 10$) time horizons.

Early life consumption and asset holdings. Another possible source of variation in saving motives is risk aversion, as highlighted in the model of Section 2. Aguiar, Bils, and Boar (2020) argue that ownership of risky assets is one way to distinguish risk preferences from time preferences. Given this observation, we construct an indicator for ownership of stocks (equities) at a long horizon ($j+10$) and use it as the dependent variable in a specification like (7). The third column of Table 8 shows that we find no evidence of a relationship between residual consumption and ownership of stocks.

In the same spirit, we check whether ownership of non-primary residences is associated with early life consumption. As shown in the last column of Table 8, there is no evidence of a relationship between ownership of non-primary residence real estate and early life consumption. These results, which hold also at shorter time horizons, indicate that heterogeneity in risk aversion is an unlikely candidate for the excess persistence in our PSID sample.

Replicating the tests using the SHIW data. Finally, Table 9 replicates the above exercises using the SHIW sample. Though non-primary real estate is not observable in the SHIW, the results for the other three dependent variables are remarkably similar to those from the PSID sample. The consumption volatility result points, again, to a role for expenditure risk heterogeneity; moreover, the wealth accumulation and stock ownership results lend support to the view that heterogeneity in time preferences is more salient than heterogeneity in risk aversion.

	Excess Persistence and Other Outcomes - SHIW			
	Consumption Volatility	Wealth Accumulation	Stock Ownership	Real Estate investment
c_t coefficient	-0.0088	-0.2597	0.0112	–
	(0.0047)	(0.1504)	(0.0118)	–
Sample Size	4286	4286	4286	–

Table 9: This table reports coefficients on c_t, measured as non-durable spending, in regressions similar to (7), but with c_{t+j} replaced by the indicated dependent variable. Real Estate Investment indicates ownership of real estate that is not the primary home.

4 A Model of Consumption Persistence over the Life Cycle

A quantitative assessment of excess persistence, and of its implications for inequality and welfare, requires a less restrictive model. We replace exponential utility with power utility (constant relative risk aversion) to generate realistic buffer-stock savings behaviors; then, consistent with the reduced form evidence in Section 3, we allow for preference heterogeneity in time preferences.[2]

As we show, permanent heterogeneity in time preferences, combined with buffer-stock behavior, can generate most of the life-cycle variation in excess persistence with one caveat: to account for excess persistence in retirement, we allow for heterogeneous preferences for end-of-life wealth. Finally, we assume that risk aversion and credit constraints are homogeneous in the population.

On the income side of the model, we generalize the unit root earnings process of the model in Section 2; specifically, we adopt a variation on the nonlinear dynamics process of Arellano, Blundell, and Bonhomme (2017) (ABB), as implemented by De Nardi, Fella, and Paz-Pardo (2020). We extend the ABB to allow for additional risk heterogeneity in earnings so to replicate the relationship between excess persistence and consumption volatility. Specifically, we take the baseline estimates by ABB as describing an average earnings process, and allow for extra variation around this average process by modeling high and low risk types whose earnings dynamics yield estimates of the ABB earnings process when aggregated.

Our calibration indicates that significant extra heterogeneity in earnings risk is needed in order to account for the relationship between expenditure volatility and excess persistence. The additional

[2]The evidence in Section 3 suggests that excess persistence is not associated with ownership of risk assets, as one would expect if heterogeneity in risk aversion played a more prominent role.

earnings variation is an important part of the information content of consumption and could not be identified in the absence of consumption data.

4.1 Model Basics

A household solves a life cycle consumption problem with random mortality up to final age T. The life cycle consists of a working stage of length T_ℓ and a retirement stage of maximum length $T - T_\ell$. During the working stage, uncertain endowments y_t are realized in each period. In retirement, income is derived from pensions and private assets.

Retirees. Retirees face mortality risk such that their probability of survival to the next period is ψ_{t+1}. In the event of death agents may value their remaining wealth according to a parsimonious function $\alpha \ln(a_{t+1})$. We posit two α-types that occur in equal proportions, one who does not value wealth at death ($\alpha = 0$), and another who does value it to a calibrated degree (α to be calibrated). Agents discount the future according to their survival probability and their idiosyncratic discount rate δ. There are two permanent δ-types which occur in equal proportions. α-type and δ-type are uncorrelated.

Given their pension income ($pens$), retirees solve the problem

$$V_t(a_t;\delta,\alpha) = \max_{c_t,a_{t+1}} \Big\{ \frac{c_t^{1-\gamma}}{1-\gamma} + \frac{1}{1+\delta}\psi_{t+1}V_{t+1}(a_{t+1};\delta,\alpha)] + (1-\psi_{t+1})\alpha\ln(a_{t+1}), \qquad (8)$$
$$s.t.\ c_t = a_t + pens - \frac{a_{t+1}}{1+r},\ a_{t+1} \geq 0 \Big\}.$$

The pension amount is common across households and set to 45% of average annual earnings.

Workers. Current earnings y_t are known but the path of future earnings is uncertain; $G_t(y_{t+1},\sigma)$ is the distribution of y_{t+1} conditional on (i) the idiosyncratic history up to period t, and (ii) a risk type σ. History dependence is necessary for ABB-style nonlinear earnings dynamics. The heterogeneity σ implies that $G_t(y_{t+1},\sigma)$ may vary across individuals even if their earnings histories are identical. Utility is discounted at a rate δ, which may also vary across households. The dynamic programming problem is:

$$V_t(a_t;\delta,\sigma,\alpha) = \max_{c_t,a_{t+1}} \Big\{ \frac{c_t^{1-\gamma}}{1-\gamma} + \frac{1}{1+\delta}E_{y_{t+1}}[V_{t+1}(a_{t+1};\delta,\sigma,\alpha)], \qquad (9)$$
$$s.t.\ c_t = a_t + y_t - \frac{a_{t+1}}{1+r},\ a_{t+1} \geq \underline{a}_t \Big\}.$$

$G_t(y_{t+1},\sigma)$ is implemented following the numerical approach developed by De Nardi, Fella, and Paz-Pardo (2020), in which simulations of the estimated non-linear earnings process are used to estimate a nonparametric Markov chain representation. To allow for types we posit that a high risk σ-type faces an earnings process where the standard deviation of possible realizations is $1+\varsigma$ times the baseline process, and a low-risk σ-type faces standard deviations $1-\varsigma$ times the baseline. By setting the

proportion of each type in the population to 0.5, we ensure that in pooled data the variance of log-earnings realizations is unchanged from the baseline, both by age and overall. The permanent δ- and α-types are independent of σ-type.

4.2 Calibration

PSID expenditure data may suffer from measurement error. Before discussing identification of key parameters, we describe how we account for the presence of (possibly non-classical) measurement error.

Measurement error. To reconcile simulated and empirical moments, we introduce random measurement error to simulated data, thereby ensuring that moments matched between model and data are consistent. We calibrate the degree of measurement error by matching the raw persistence of consumption across several time horizons. For example, in a simple regression of c_{t+2} on c_t in our PSID sample, the coefficient value of 0.716 is considerably lower than a reasonable calibration of the model would generate. We therefore set the standard deviation of measurement error so that the implied attenuation bias leads to a match between model-generated and PSID data coefficients at that horizon. Persistence in measurement error is another potential concern, i.e. the possibility that households misreport in a similar manner over time. One can show that measurement error of this nature would bias the coefficient upwards to a degree that increases with the time horizon. Thus, a permanent component of measurement error can be calibrated to match the flatness of the PSID consumption persistence profile. We find that a time-varying standard deviation of (log additive) measurement error of about 0.25, combined with a permanent measurement error with standard deviation of about 0.1, replicates the profile of raw consumption persistence in the PSID quite well.

4.2.1 Identification of Model Parameters

Next, we overview the identification of model parameters. There are five key features of the model to be identified: (1) σ-heterogeneity, (2) δ-heterogeneity, (3) α-heterogeneity, (4) the common γ parameter, and (5) the common credit limit $\underline{a}$.

Heterogeneity: Time preferences vs earnings risk. The simple model of Section 2 illustrates the identification challenge of distinguishing time-preference heterogeneity from earnings-risk heterogeneity, as both would affect excess-persistence similarly. This is a very consequential distinction if one wants to answer questions about the welfare implications of consumption inequality.

The key empirical moment that allows to separately identify these features is the correlation between residual spending at time t and the volatility of future spending (in periods $t+2$ through $t+10$). This relationship could be positive or negative depending on which of two forces dominates. If residual spending is determined entirely by δ-heterogeneity (time preferences), those who spend less than expected will exhibit a lower volatility of future spending as they accumulate a larger stock of precau-

tionary wealth. Hence, δ-heterogeneity alone would lead to a positive relationship between residual spending and realized volatility. However, if σ-heterogeneity is the driving force, we should see lower spending among those who anticipate larger earnings volatility in the future, likely resulting in a muted or negative relationship between residual spending and realized consumption volatility. Hence, the observed negative relationship tells us that σ-heterogeneity plays a large role and the parameter ς can be identified by matching the model relationship between residual spending and realized consumption volatility to the data.

Though the extent of σ-heterogeneity implied above is reasonably large, it is not nearly enough to explain the measured excess persistence (this is true even after accounting for persistent measurement error). Thus, the degree or δ-heterogeneity is identified as the time-preferences variation necessary to induce enough excess persistence in the model to match the data.

End-of-life preferences. The remaining parameters are identified from variation in excess persistence over the life cycle and over different time horizons. In particular, α-heterogeneity is specified so that excess persistence during retirement is as strong as in data. Recall from Figure 1 that excess persistence follows a hump-shape over the life cycle; yet, it does not disappear during retirement years. In a model with no α-heterogeneity, excess persistence would be counter-factually low in retirement because of the dissolving earnings risk. The α-heterogeneity implies that end-of-life wealth becomes more salient for the spending decisions of households at older ages.

Utility curvature and borrowing limit. To understand how the utility curvature (γ) and borrowing limit ($\underline{a}$) are identified from life cycle variation in excess persistence, one must consider how buffer stock savings behavior relates to persistence in the model. Households engaged in buffer-stock savings behavior cannot accumulate unbounded precautionary wealth. Rather, they save (or dis-save) until their buffer stock target is reached. Although permanent heterogeneity would lead to constant λ_{t+k} terms under exponential utility, in the buffer-stock model the forces that cause excess persistence fade out over time as buffer stocks are attained. Intuitively, two households may have different buffer-stock targets due to time-preference heterogeneity but, after they attain those targets, their different levels of wealth shall reveal their type. Given that the heterogeneous decay in savings motives crucial for the intensity of excess persistence, the speed of adjustment towards buffer stock targets is crucial for replicating how excess persistence varies over different time horizons in the data.

Jappelli and Pistaferri (2024) show that the inter-temporal elasticity of substitution ($1/\gamma$) is an important determinant of the speed of adjustment of buffer-stock wealth. We set that parameter to reflect the rate of decline of excess persistence as the time horizon j increases. As argued above, the overall level of excess persistence is determined by δ-heterogeneity in time-preferences, whereas variation over time horizons is determined by the homogenous curvature parameter γ.

The last parameter to identify is the (homogeneous) credit constraint, $\underline{a}$. This parameter has the largest impact on excess persistence among very young households who are the most likely to hit the credit constraint. There is a tendency in the model for excess persistence among the very young to

be low when the credit constraint is very tight. This is because two credit constrained households with the same income will make the same consumption choices, even if their time-preferences are different. By loosening the borrowing limit these households become free to make persistently different consumption choices, increasing the degree of excess persistence. Identification of the credit constraint is thus achieved by loosening the borrowing limit until the degree of excess persistence in the model matches the data.

4.2.2 Parameters and Model Fit

Table 10 reports the calibrated parameter values and the main target moment for identification for each parameter. Additionally, Figure 2 illustrates how the model fits the PSID excess persistence results (using non-durable spending), and Table 11 reports replication of regressions with future consumption volatility and wealth as the dependent variables.

Parameter	Calibrated Values	Target Moment
δ	$\{.036, .129\}$	Excess Persistence Level (Fig 2)
γ	2.4	Excess Persistence j-Variation (Fig 2)
ς	0.6	Volatility Regression (Table 11)
α	2.1	Retired Excess Persistence (Fig 2)
$\underline{a}$	$-.62 \times E[y]$	Young Excess Persistence (Fig 2)

Table 10: Calibration Summary.

	Volatility and Wealth Regressions			
	Consumption Volatility (Data)	Wealth Accumulation (Data)	Consumption Volatility (Model)	Wealth Accumulation (Model)
c_t coefficient	-0.0115	-0.1836	-0.0114	-0.2182
	(0.0044)	(0.0778)	–	–

Table 11

A substantial degree of δ-heterogeneity is needed to replicated the overall degree of excess persistence measured in the data. To see how important this dimension of heterogeneity is for explaining the data, Figure 3 shows how excess persistence would evolve over the life cycle if all households had the average of the calibrated discount factors. The very large generalized decreases in excess persistence when δ is homogeneous show how important matching excess persistence is for calibrating discount rate heterogeneity.

Removing income risk heterogeneity also reduces excess persistence, as Figure 4 shows, though not to the same degree as removing discount rate heterogeneity. This raises again the question of how to tell which feature is responsible for observed excess persistence. As argued above, we turn to the relationship between unexplained spending and realized consumption volatility to sort this out. In

Figure 2: Each line is the $j \in \{2, 4, 6, 8, 10\}$ sequence for a particular age. The base ages are 30, 40, ... 70.

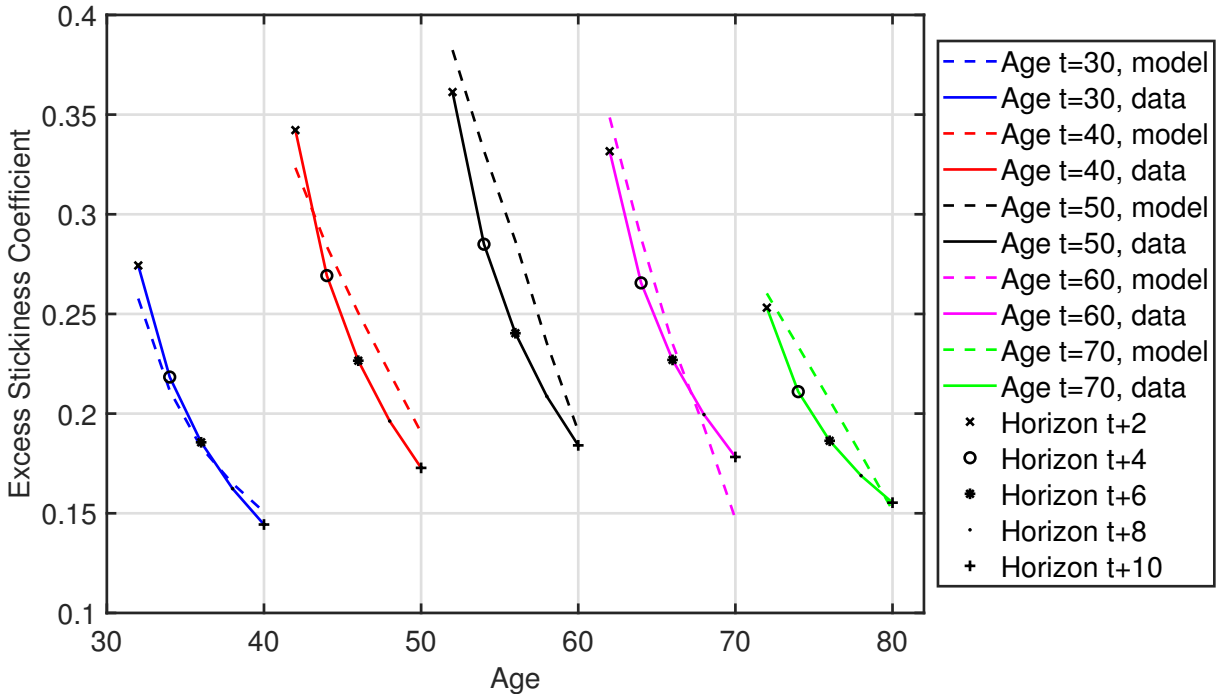
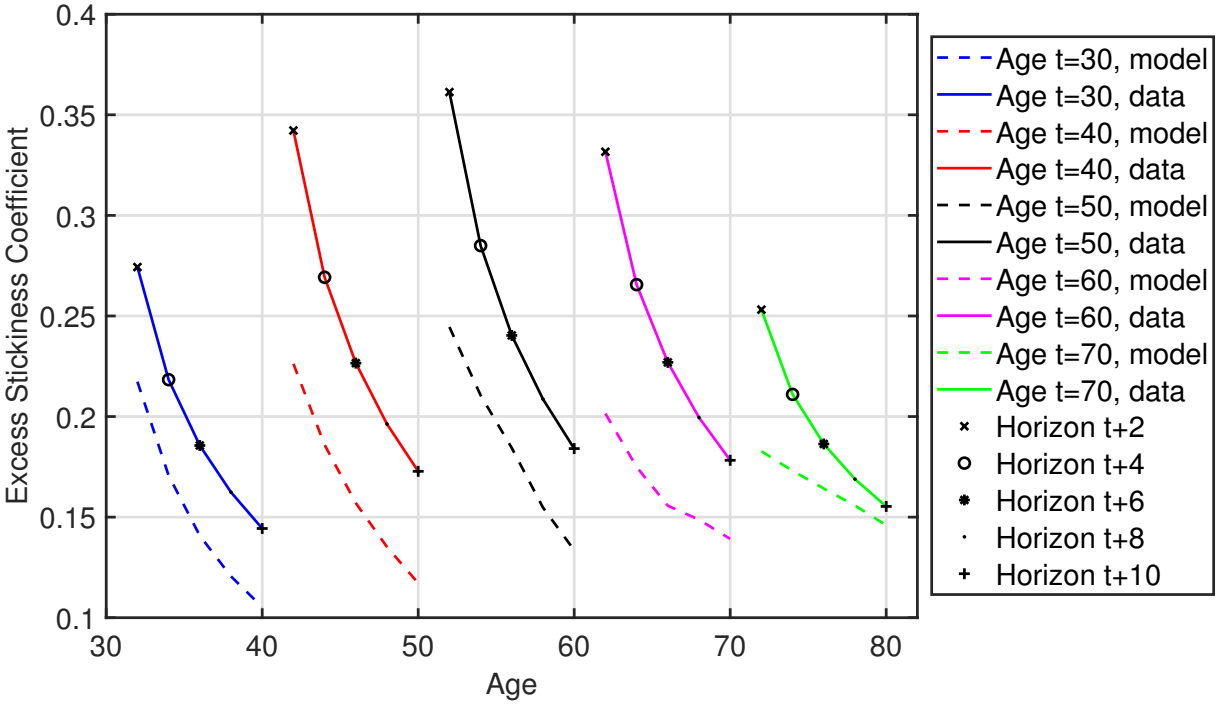
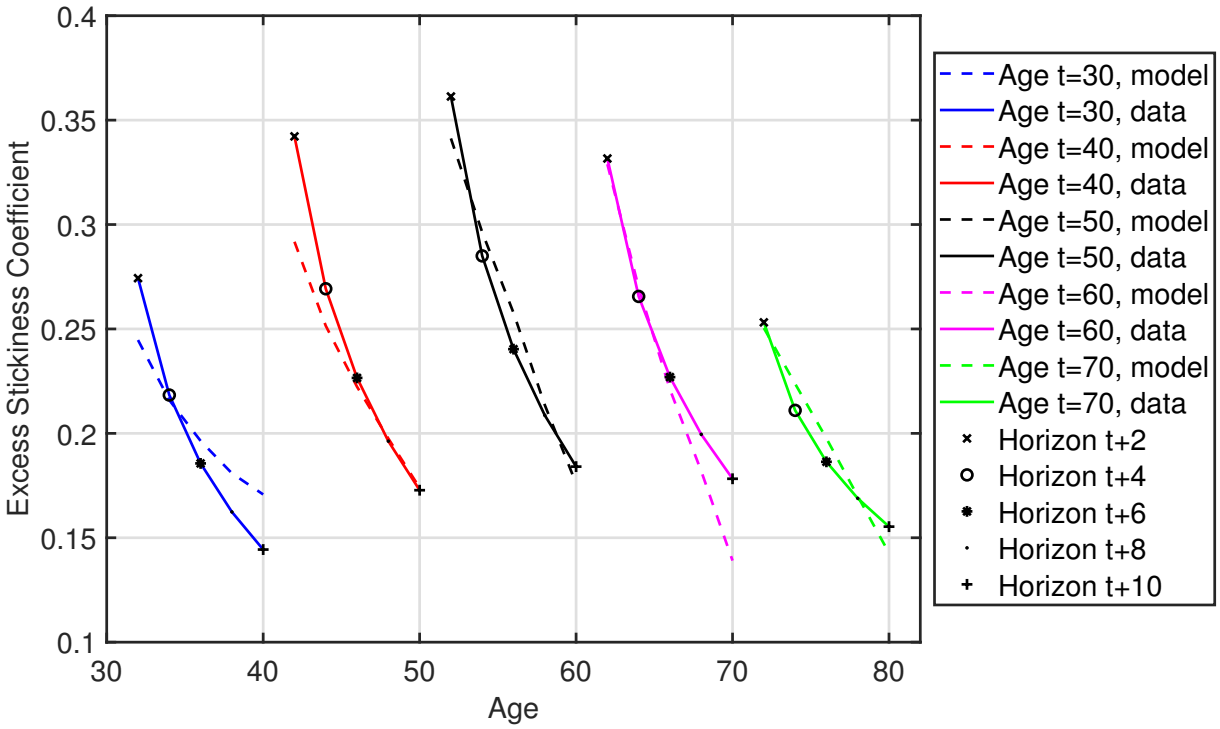


Figure 3: Each line is the $j \in \{2, 4, 6, 8, 10\}$ sequence for a particular age. The base ages are 30, 40, ... 70.



the absence of σ-heterogeneity, the coefficient in the consumption volatility regression turns positive (+0.003) meaning that high savers endogenously end up with less volatile consumption realizations. On the other hand, when δ-heterogeneity is removed, the negative coefficient becomes more pronounced, increasing in absolute value to -0.032.

Figure 4: Excess persistence with σ-heterogeneity removed.



The remaining parameters also generate variation in excess persistence as described in section 4.2.1. We provide plots to illustrate this variation in Appendix A. For example, eliminating α-heterogeneity (Figure 7) reduces excess persistence primarily for retired households, while increasing the elasticity of intertemporal substitution (reducing γ, Figure 9) results in much greater variation in excess persistence over horizons j, particularly for middle-aged households. Cutting the credit limit by half, from about 62% of average to 31% (Figure 8), reduces the excess persistence of young households because a greater prevalence of constrained choices reduces persistent heterogeneity in consumption decision rules. Lastly, Figure 10 shows that excess persistence β coefficients are reduced by about 0.05-0.1 when the permanent component of measurement error is removed, thus illustrating the non-trivial role of persistent measurement error in the data.

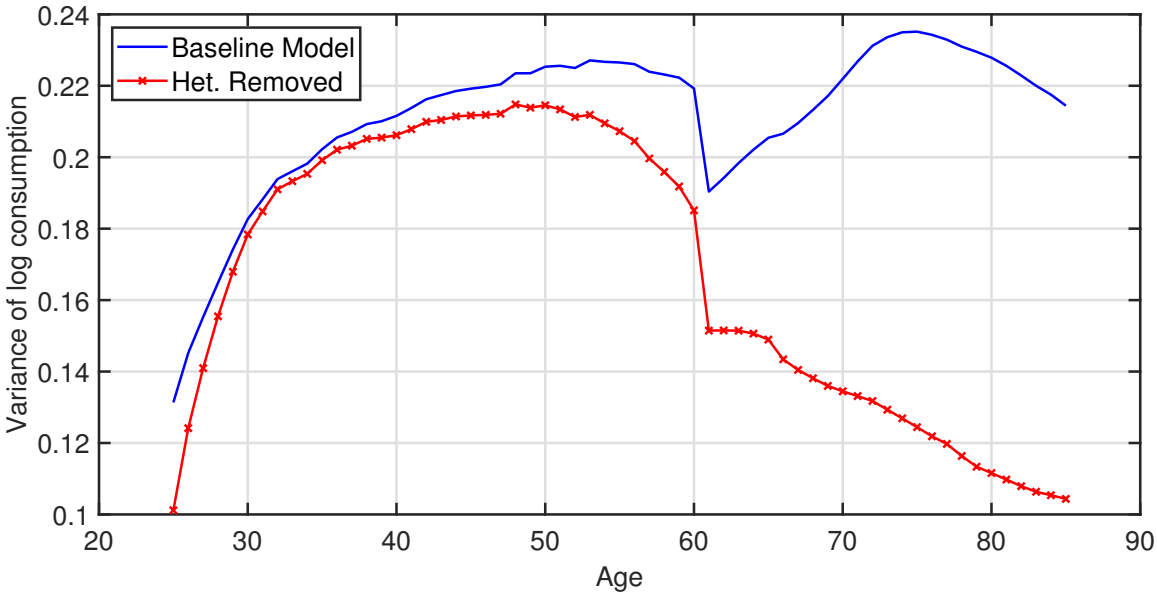
5 Consumption Dispersion and Welfare

It is instructive to begin with an examination of how consumption inequality varies over the life cycle, and how the different sources of heterogeneity in our model contribute to this. To this end, Figure 5

plots two versions of the variance of log-consumption at different ages.

Heterogeneity and consumption inequality over the life cycle. The baseline model generates realistic levels of consumption inequality, as well as the well-known rise in the variance of log-consumption during working ages. In retirement, the model generates a drop in consumption inequality as the pension age is reached, and then a recovery as life progresses further.

Figure 5: Variance of log-consumption over the life cycle, for the baseline model and with heterogeneity fixed-effects removed.



Next, we plot of the variance of log-consumption after removing variation due to fixed effects for each of σ, δ and α heterogeneity. We make two observations: (i) all those sources of heterogeneity play an important role for inequality; (ii) they have an uneven influence at different stages of the life cycle. For the young, about 1/4 of consumption variation is accounted for by ex-ante heterogeneity (σ, δ and α); however, for workers in their 30s these parameters contribute relatively little. For the remainder of the life cycle there is a growing contribution to overall consumption inequality by ex-ante heterogeneity, with more than half of the variance of log-consumption among the oldest being generated by those parameters.

Changing saving patterns explain why the influence of ex-ante heterogeneity first narrows then grows. For the young, those with either low-δ or high-σ save more than their counterparts, leading to large gaps in consumption. As their wealth expands, the spending of the low-δ or high-σ households catches up, leading to a stage of the life cycle where ex-ante heterogeneity accounts for a relatively little share of consumption inequality. Later in life, the richer households begin to consume more, leading to a growing share of consumption inequality being explained by ex-ante heterogeneity. During retirement, we observe that consumption inequality monotonically decreases after accounting for ex-

ante heterogeneity; thus, differences in savings motives are responsible for the post-retirement increase of inequality that we see in the baseline model.

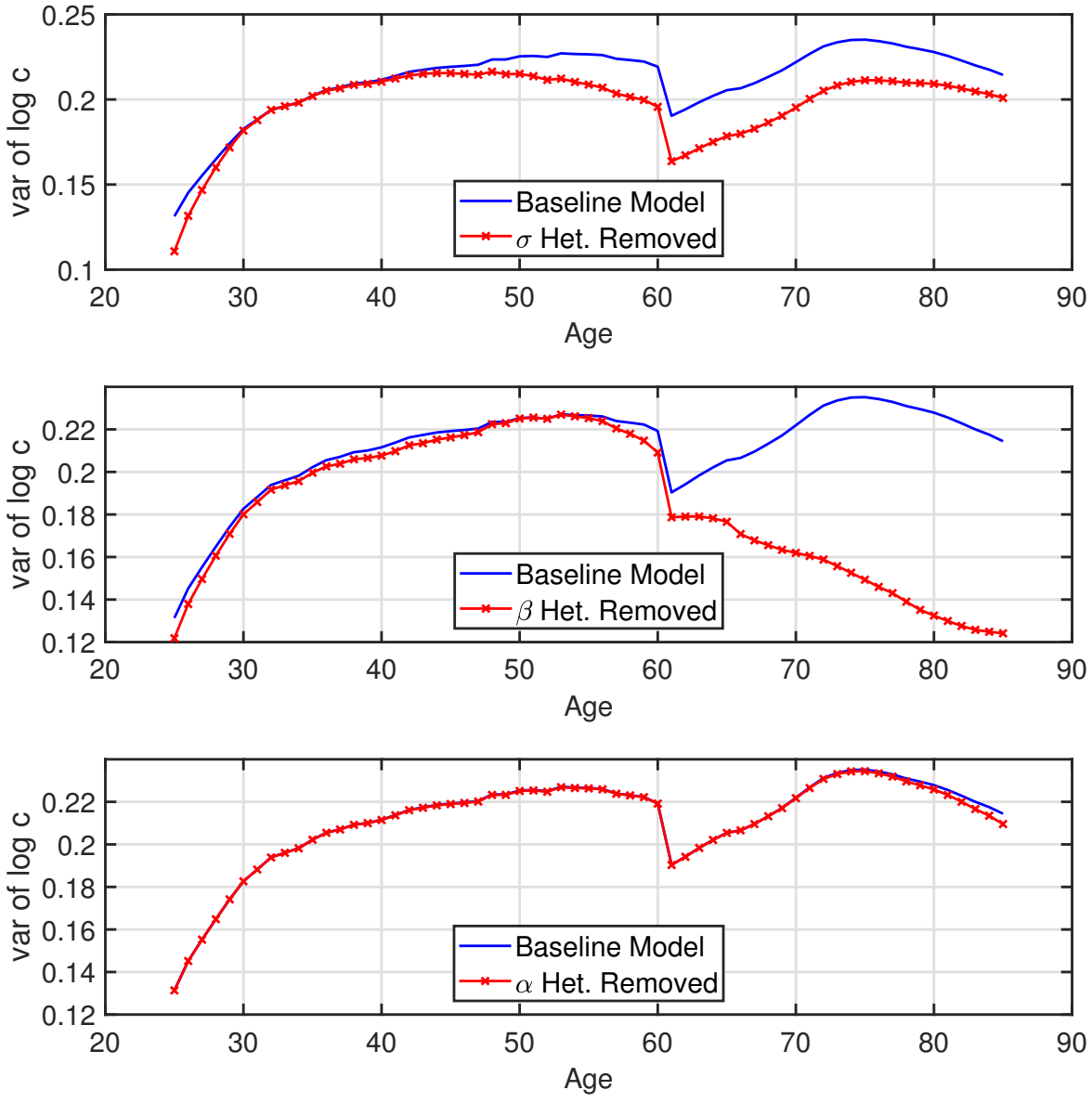
Next, in Figure 6, we separately remove the variation due to each source of ex-ante heterogeneity. The top panel shows that much of the contribution of ex-ante heterogeneity to early life consumption inequality occurs through σ-heterogeneity; moreover, the subsiding influence of ex-ante heterogeneity in the 30s is due to consumption catch-up by high-risk households due to their greater savings. Larger accumulated savings eventually lead to higher consumption among high-σ households, which persists through retirement.

When variation due to β heterogeneity is removed in the middle panel of Figure 6 we observe a much longer period of time before the consumption of patient households catches up to impatient households, which results the point where δ-heterogeneity does contribute to overall inequality being around age 50. For the periods up to age 50 patient households consume less, and after age 50 they consume more, leading to a role for time-preference in overall consumption inequality at those ages. This becomes particularly pronounced in retirement when large differences in accumulated resources tend to spent. We can see from the graph that the rise of consumption inequality during retirement in the baseline model is due to time preference-heterogeneity. More patient households delay spending their accumulated wealth longer than impatient households, leading them to have proportionally much greater consumption in their 70s. Heterogeneity in preferences for end-of-life wealth have relatively little effect on consumption inequality, except very late in life when they become more salient.

Consumption inequality vs welfare inequality. The overall variance of log-consumption in our model economy is 0.2257; however, the life-cycle analysis above illustrates that a considerable part of that may be due to welfare irrelevant preference heterogeneity. To be clear, not all ex-ante heterogeneity is welfare-irrelevant. In particular, consumption inequality due to ex-ante heterogeneity in the degree of risk should is relevant for welfare because equalization of consumption between high and low risk types can increase average utility in the economy (improve utilitarian welfare). However, equalization of consumption between high and low patience consumers would not be beneficial from a utilitarian-welfare perspective because these consumers optimally choose to allocate resources differently over their life cycle.

To gauge the role of welfare irrelevant preferences we remove variation due to δ-type and α-type fixed-effects, fully interacted with age fixed-effects. The interaction with age fixed-effects is to account for the fact that some variation in wealth is the results of preference heterogeneity, and we want to capture this variation. When variation due to these fixed effects is removed, the variance of log-consumption drops to 0.1871, indicating that about 17% of cross-sectional consumption inequality is purely due to preference variation. If we further exclude variation explained by σ-type fixed effects (interacted with age fixed-effects) the variance of log-consumption falls further to 0.1737, showing that a further 6% of consumption variation is due to this factor. For context, 6% is also the proportion of the remaining cross-sectional spending variation that is explained by age-effects, thus σ-heterogeneity can still be interpreted as quite salient despite the small percentage.

Figure 6: Variance of log-consumption over the life cycle, with variation due to each source of ex-ante heterogeneity removed separately.



Of the variation in consumption that is orthogonal to permanent income, about 1/4 is due to a welfare-relevant factor, and 3/4 is due to welfare-irrelevant heterogeneity. The next section will then ask whether adding this information to that contained in permanent-income improves or impairs measurement of welfare inequality.

6 Conclusion

The measurement, and origins, of inequality are of first-order importance for the design of policies. The increasing availability of household-level expenditure data has generated much interest in using the distribution of consumption as a measure of cross-sectional welfare differences. We examine the information content of expenditure data and how to interpret them in relation to more standard data on earnings, income and wealth. Specifically, we ask: what information do household expenditures convey? How much of this information is above and beyond what we can already learn from the distributions of earnings, income and other measures of permanent-income?

To distinguish between different sources of consumption inequality, we introduce the notion of *excess persistence* of consumption. This refers to persistence that is above and beyond what can be accounted for through wealth and income histories. Excess persistence arises when households engage in different consumption choices over a given time horizon and in the face of measurably similar circumstances. Possible explanations for such discrepancies include heterogeneity in preferences or in the distribution of risks and constraints.

We use a life cycle model to show that the excess persistence is related to the volatility of consumption. Then, we leverage this relationship to parse consumption inequality into components due to resources, preferences and risk. We examine two household panels, the U.S. PSID and the Italian SHIW, and document similar patterns of excess persistence in both.

A quantitative assessment illustrates how much of the variation in idiosyncratic preferences, risk, and resources is implied by our estimates of excess persistence. Roughly 77% of consumption inequality can be reconciled through permanent-income measures based on income, wealth and other household data; of the remaining consumption variation, about 17% can be attributed to heterogeneous preferences and 6% is due to private information about idiosyncratic expenditure risk that cannot be measured from other data. The latter is about as large as the consumption variation due to age. The importance of preference heterogeneity and private information suggests that caution is needed when using consumption inequality as a measure of the efficacy of redistributive policies. Information from income and wealth data might be preferable when assessing policy impacts on heterogeneous populations.

References

Aguiar, M. A., Bils, M., and Boar, C. (2020). 'Who are the Hand-to-Mouth?' Tech. rep., National Bureau of Economic Research.

Arellano, M., Blundell, R., and Bonhomme, S. (2017). 'Earnings and consumption dynamics: a nonlinear panel data framework.' *Econometrica*, vol. 85(3):pages 693–734.

Blundell, R., Pistaferri, L., and Preston, I. (2008). 'Consumption Inequality and Partial Insurance.' *The American Economic Review*, vol. 98(5):pages 1887–1921.

Blundell, R. and Preston, I. (1998). 'Consumption inequality and income uncertainty.' *The Quarterly Journal of Economics*, vol. 113(2):pages 603–640.

Caballero, R. J. (1990). 'Consumption puzzles and precautionary savings.' *Journal of monetary economics*, vol. 25(1):pages 113–136.

Caballero, R. J. (1991). 'Earnings uncertainty and aggregate wealth accumulation.' *The American Economic Review*, pages 859–871.

Chetty, R. and Szeidl, A. (2007). 'Consumption commitments and risk preferences.' *The Quarterly Journal of Economics*, vol. 122(2):pages 831–877.

Chetty, R. and Szeidl, A. (2016). 'Consumption commitments and habit formation.' *Econometrica*, vol. 84(2):pages 855–890.

Daminato, C. and Pistaferri, L. (2023). 'Heterogeneity in returns to wealth and consumption inequality.' Tech. rep., CEPR, DP19093.

De Nardi, M., Fella, G., and Paz-Pardo, G. (2020). 'Nonlinear household earnings dynamics, self-insurance, and welfare.' *Journal of the European Economic Association*, vol. 18(2):pages 890–926.

Insolera, N. E., Simmert, B. A., and Johnson, D. S. (2021). 'An overview of data comparisons between psid and other us household surveys.' *Technical Series Paper*, pages 21–02.

Jappelli, T. and Pistaferri, L. (2006). 'Intertemporal choice and consumption mobility.' *Journal of the European Economic Association*, vol. 4(1):pages 75–115.

Jappelli, T. and Pistaferri, L. (2024). 'Permanent Income Shocks, Target Wealth, and the Wealth Gap.' *American Economic Journal: Macroeconomics*.

A Identification Plots

Figure 7: Excess persistence with α heterogeneity removed.

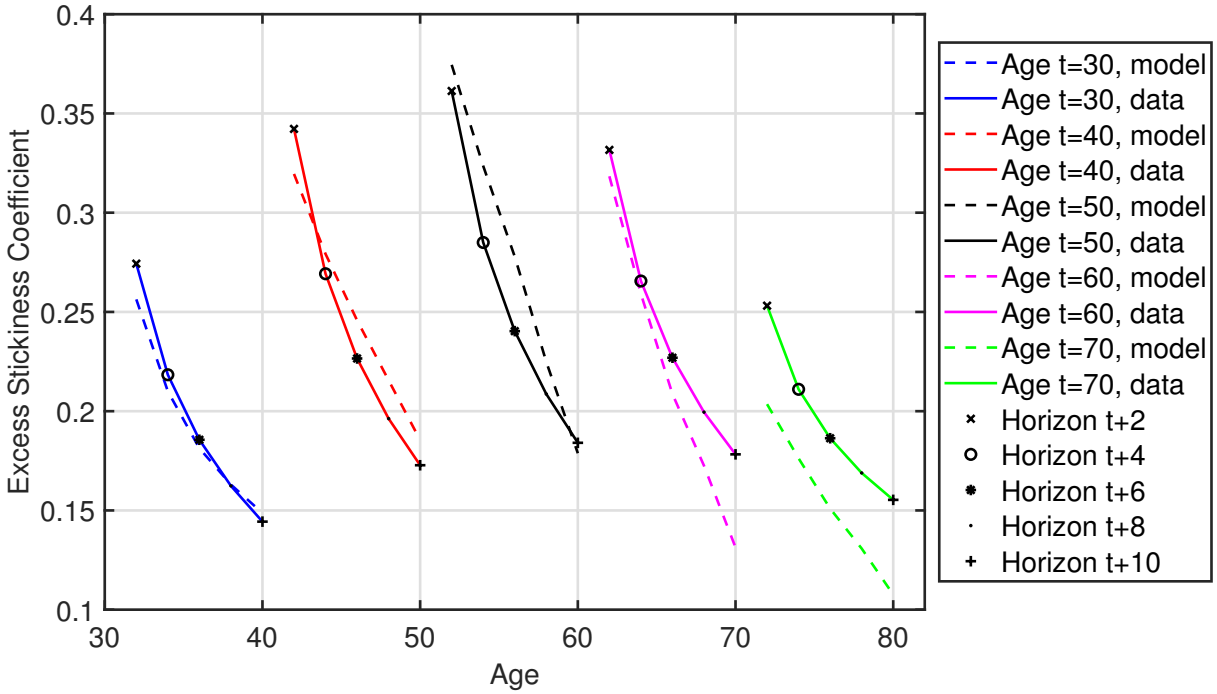


Figure 8: Excess persistence with borrowing limit halved.

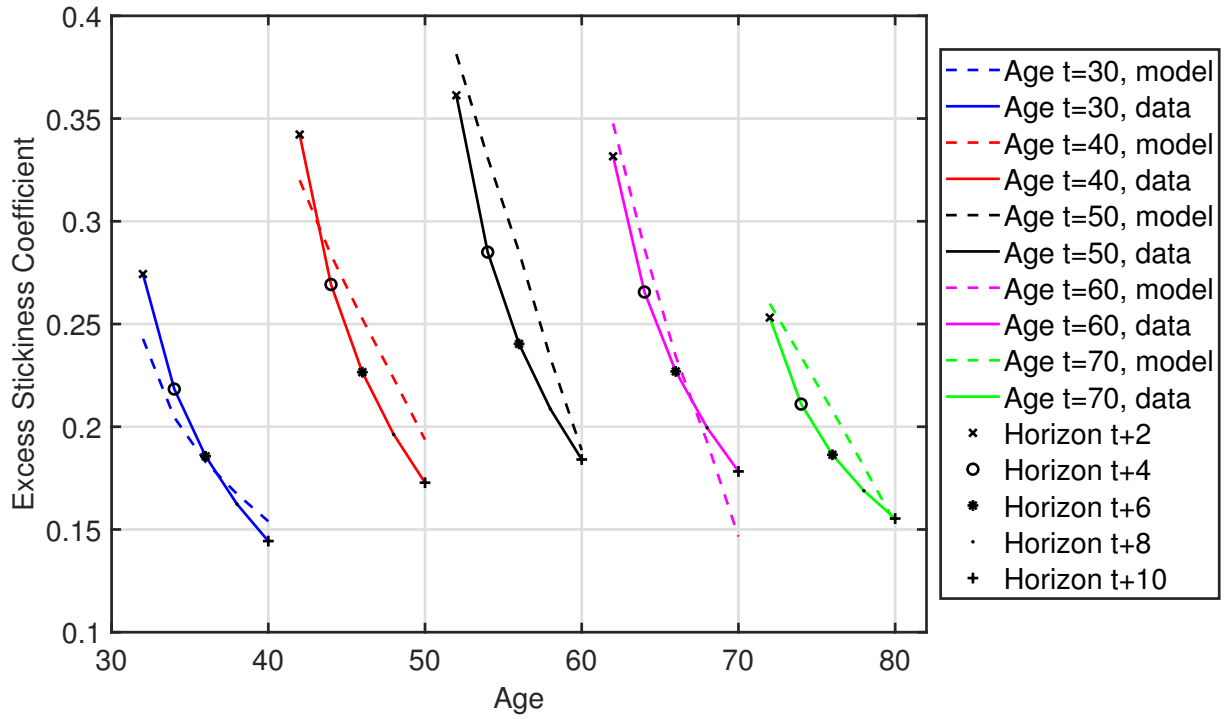


Figure 9: Excess persistence with $\gamma = 2$.

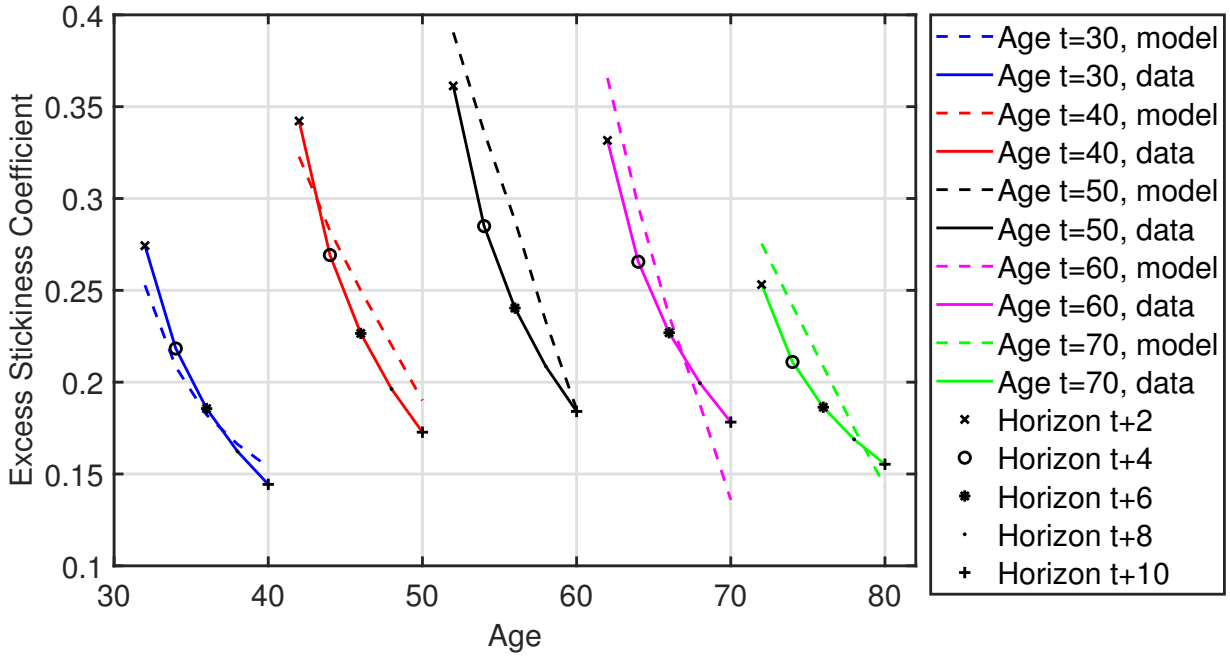


Figure 10: Excess persistence with permanent measurement error component removed.

